PHYLOGENETIC AND BIOINFORMATIC ANALYSIS OF GENES RELATED TO IMMUNITY AND RETROVIRAL ELEMENTS IN *NASALIS LARVATUS* GENOME

ASWINI LEELA LOGANATHAN

THESIS SUBMITTED IN FULFILLMENT FOR THE
DEGREE OF MASTER OF SCIENCE

BIOTECHNOLOGY RESEARCH INSTITUTE

UNIVERSITI MALAYSIA SABAH

2014

UNIVERSITI MALAYSIA SABAH

BORANG PENGESAHAN STATUS THESIS

JUDUL : **PHYLOGENETIC AND BIOINFORMATIC ANALYSIS OF GENES RELATED TO IMMUNITY AND RETROVIRAL ELEMENTS IN *NASALIS LARVATUS* GENOME**

IJAZAH. : **SARJANA SAINS**

Saya Aswini Leela Loganathan, Sesi Pengajian 2012-2014, mengaku membenarkan tesis Sarjana ini disimpan di Perpustakaan Universiti Malaysia Sabah dengan syarat-syarat kegunaan seperti berikut:-

1. Tesis ini adalah hak milik Universiti Malaysia Sabah
2. Perpustakaan Universiti Malaysia Sabah dibenarkan membuat Salinan untuk tujuan pengajian sahaja.
3. Perpustakaan dibenarkan membuat Salinan tesis ini sebagai bahan pertukaran antara institusi pengajian tinggi.

☐ SULIT (Mengandungi maklumat yang berdarjah keselamatan atau kepentingan Malaysia seperti yang termaktub di dalam AKTA RAHSIA RASMI 1972)

☐ TERHAD (Mengandungi maklumat TERHAD yang ditentukan oleh organisasi/ badan di mana penyelidikan dijalankan

☑ TIDAK TERHAD

Disahkan oleh,

NURULAIN BINTI ISMAIL
LIBRARIAN
UNIVERSITI MALAYSIA SABAH

ASWINI LEELA LOGANATHAN

(PERPUSTAKAWAN)

ALAMAT: No2, Jalan Anggerik 15
Anggerik Villa, Taman Kulai Utama,
81000, Kulai, Johor

Tarikh: 7 Disember 2014

(PROF. MADYA. DR. VIJAY KUMAR)

DECLARATION

I hereby declare that the material in this thesis is my own except for quotations, excerpts, equations, summaries and references, which have been duly acknowledged.

7th December 2014

Aswini Leela Loganathan
MZ1211011T

CERTIFICATION

NAME : **ASWINI LEELA LOGANATHAN**

MATRIC NO. : **MZ1211011T**

TITLE : **PHYLOGENETIC AND BIOINFORMATIC ANALYSIS OF GENES RELATED TO IMMUNITY AND RETROVIRAL ELEMENTS IN *NASALIS LARVATUS* GENOME**

DEGREE : **MASTER OF SCIENCE IN BIOTECHNOLOGY (MOLECULAR GENETIC)**

VIVA DATE : **5TH DECEMBER 2014**

DECLARED BY

1. SUPERVISOR COMMITTEE

Signature

Assoc. Prof. Dr. Vijay Kumar

Dr. Kenneth Francis Rodrigues

Dr. Christopher Voo Lok Yung

ACKNOWLEDGEMENTS

To the Almighty God who has granted me all these graces to fulfill this work and who supported me and blessed me by His power and His mercy in all my life. To Him I extend my heartfelt thanks. Many institutions and individuals were responsible for the crystallization of this humble work, whose associations and encouragement have contributed to the accomplishment of the present thesis, and I would like to pay tribute to all of them. I specially wish to express my sincere thanks and gratitude to my supervisor, Dr. Kenneth Francis Rodrigues for his kindness, admirable supervision, direct guidance, generous considerations and valuable support during my research under him. Not forgetting his fruitful discussions, constructive advises, and particularly for sharing his expertise in molecular genetics and revision of this thesis and giving me the opportunity to do my research in University Malaysia Sabah (UMS). I would also like to express my gratefulness to my co - supervisors of my thesis, Assoc. Prof. Dr. Vijay Kumar and Dr. Christopher Voo Lok Yung, for their motivations, and for teaching me very precious knowledge about molecular genetics and providing basic bioinformatic knowledge and offering me such a kind of opportunity regarding my research and their continuous encouragement and kind advises. My deepest thanks to all the staff members and colleagues of the Biotechnology Research Institute for their kindness and guidance and for always helping me with the research. My gratitude to University Malaysia Sabah (UMS) for giving me this opportunity to further increase my knowledge and experience by helping me to do my research in this University. Lastly I would like to extend my special thanks to my family and friends for their never ending love, support and encouragement.

ASWINI LEELA LOGANATHAN
7TH DECEMBER 2014

ABSTRACT

The Proboscis monkey (*Nasalis larvatus*) also known locally as *Monyet Belanda / Bayau* are endemic to the island of Borneo. Unfortunately the *N. larvatus* habitat loss has resulted in severe decline in endemic population and listed in Appendix I by the Convention on International Trade in Endangered Species (CITES). This research is divided into two parts. The first part of this research focused on the Mitochondrial DNA (mtDNA) of the *N. larvatus*. At present, there are only 2 data available *for N. larvatus* mtDNA and this research served as a benchmark for validating those data. Furthermore the importance of this study is to amplify the mitochondrial genome of the species *Nasalis larvatus* from two regions which is Sabah and Sarawak, which is then utilized for sequencing and ultimately comparative mitochondrial genomics with other Primates. Second part of this research focused on characterizing the genes linked to immunity and host defence which are the Major histocompatibility complex (MHC), Defensins and Retroviral elements present in the *N. larvatus* genome. In order to achieve this, the polymerase chain reaction (PCR) primers were designed from a total of 31 immune genes which were MHC, Defensins and Retroviruses of Non human primate genes obtained from EST. Cloning was carried out using CloneJET PCR cloning kit (Thermo Scientific). The plasmid was purified using Gene Jet Plasmid Mini Prep Kit (Thermo Scientific). The plasmid was then sequenced using BigDye Terminator v3.1 cycle sequencing kit chemistry. All novel immunology and retrovirus sequences have been deposited in GenBank (accession no: KF279631-KF279640, KP064501-KP064501 and KM889667). From the study, the mitochondrial genome *of N. larvatus* from Sabah and Sarawak is 16,574 bp and 16,573 bp long, respectively. Selection at the amino acid level was measured as the rates of nonsynonymous (d_N) and synonymous (d_S) substitutions per codon site. From the data, mtDNA genes underwent neutral selection. For the MHC data, there were no positive selection detected while for defensin and retrovirus data, it shows that some primates underwent neutral selection and some underwent positive selection. The phylogenetic tree was created using the Neighbor-Joining and maximum likelihood method. The MHC data shows that the tree topology is orthologous with the primates. As for the defensins, phylogenetic tree results show it is highly paralogous with *Macaca mulatta* while as for the Retroviruses the phylogenetic tree shows orthologous divergent. Because the two different methods used, give trees with the same topology, the phylogenetic relationships derived from these sequence comparisons appear reliable. In this study, the aim is not to get the complete coverage of the retroviral and immunology genes but to compare the sequences to *N. larvatus* to see if they are homologous.

ABSTRAK

ANALISIS FILOGENETIK DAN BIOINFOMATIK GENE GENE YANG BERKAITAN DENGAN IMUNITI DAN ELEMEN RETROVIRAL DI DALAM GENOM NASALIS LARVATUS.

Nasalis larvatus (Monyet Belanda atau Bayau) adalah endemik pada kepulauan Borneo. Malangnya N. larvatus ini disenaraikan sebagai lampiran I oleh Convention on International Trade in Endangered Species (CITES). Kajian ini dibahagikan kepada dua bahagian. Bahagian pertama kajian ini akan memberi tumpuan kepada DNA mitokondrium (mtDNA) daripada N. larvatus. Maklumat yang diperolehi akan membantu dalam mentafsir order filogenetik dalam Kingdom Animalia. Pada masa ini, hanya terdapat 2 data berkaitan mtDNA N. larvatus dan penyelidikan ini boleh dijadikan sebagai tanda aras bagi mengesahkan data tersebut. Bahagian kedua kajian ini akan menumpukan pada pengkarektoran gen yang dikaitkan dengan imuniti dan Major histocompatibility complex (MHC), Defensins dan unsur-unsur Retroviral yang hadir dalam genom N. larvatus. Gen-gen ini mempunyai beberapa peranan yang unik selain memainkan peranan penting dalam sistem pertahanan badan. Bagi mencapai matlamat ini, Rantaian tindak balas (PCR) polymerase direka dari sejumlah 31 gen imun yang terdiri dari MHC, Defensins dan retrovirus gen manusia dan primat yang diperolehi dari EST. Kloning dibuat menggunakan kit pengklonan CloneJET PCR (Thermo saintifik). Plasmid dipurifikasi dengan menggunakan kit Gene Jet Plasmid Mini Prep (Thermo saintifik). Kemudian, plasmid yang dipurifikasi itu disequenced menggunakan kit kimia BigDye Terminator v3.1 kitaran penjujukan. Kesemua sequence imunologi dan retrovirus novel ini telah dipamerkan di GenBank (Nombor aksesi: KF279631 - KF279640 , KP064501-KP064501 dan KM889667). Dari kajian ini , genom mitokondria N. larvatus dari Sabah dan Sarawak adalah 16574 dan 16573 bp masing-masing. Pemilihan di peringkat asid amino telah diukur sebagai kadar substitusi nonsynonymous (dN) dan synonymous (dS) pada setiap codon. Dari data itu, gen mtDNA telah menjalani pemilihan neutral. Untuk data MHC, tidak ada pilihan positif dikesan manakala bagi data defensin dan retrovirus, ia menunjukkan bahawa segelintir primat menjalani pemilihan neutral dan segelintir menjalani pemilihan positif. Pokok filogenetik telah dicipta menggunakan kaedah Neighbor-joining dan Maximum likelihood. Data menunjukkan bahawa, bagi data MHC, topologi pokok adalah ortologos dengan primat . Bagi data defensins, keputusan pokok filogenetik menunjukkan ia adalah paralogous dengan Macaca mulatta manakala untuk retrovirus pokok filogenetik menunjukkan ortologos dengan primat. Oleh kerana kedua-dua kaedah yang berbeza digunakan, memberikan pokok dengan topologi yang sama, hubungan filogenetik yang diperolehi daripada perbandingan urutan ini adalah dipercayai . Dalam kajian ini, matlamatnya bukan untuk mendapatkan liputan yang lengkap gen retroviral dan imunologi tetapi untuk membandingkannya dengan N. larvatus bagi melihat jika ianya adalah homolog.

TABLE OF CONTENT

CHAPTER 1: INTRODUCTION

CHAPTER 2: LITERATURE REVIEW

CHAPTER 3: MATERIALS AND METHODS

CHAPTER 4: RESULTS

CHAPTER 5: DISCUSSION

LIST OF TABLES

LIST OF PHOTOGRAPHS

LIST OF ABBREVIATIONS

mtDNA	-	Mitochondrial DNA
DNA	-	Deoxyribonucleic acid
MHC	-	Major Histocompatibility Complex
TLR	-	Toll Like Receptor
OWM	-	Old World Monkey
DARC	-	Duffy Antigen Receptor for Chemokines
ERV	-	Endogenous Retroviruses
HERV	-	Human Endogenous Retrovirus
TRIM	-	Tripartite Motif
NJ	-	Neighbour-Joining
ML	-	Maximum Likelihood
%	-	Percent
>	-	More than
<	-	Less Than
α	-	Alpha
β	-	Beta
θ	-	Theta
°C	-	Degree Celsius
BLAST	-	Basic alignment search tool
bp	-	Base pair
dNTP	-	Deoxynucleoside-5′-triphophase
dH_2O	-	Distilled Water

EDTA - Ethylenediaminetetra-acetate

e-value - Expected value

mm - Milimeter

kb - Kilo base

mM - Milimolar

ml - Mililitre

$MgCl_2$ - Magnesium Chloride

µ - Micro

M - Molar

µL - Microlitre

µg - Microgram

µmol - Micromole

nm - Nanometer

PCR - Polymerase Chain Reaction

RNA - Ribinucleic Acid

TBE - Tris-Boric Acid- EDTA

U - Unit

V - Volt

ti - Transition

tv - Transversion

dS - Synonymous

dN - Non-Synonymous

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Background

Sabah, (72,000km^2) is a Malaysian state which occupies the north-eastern tip of Borneo Island in Southeast Asia. It is situated within one of the richest bio-geographical regions equally in terms of species and habitat richness. Its non-human primate fauna in particular varies greatly. Thirteen species of non-human primate have been known to live on Borneo Island, and 10 species are reputable in Sabah. Amongst the most outstanding non-human primate species is the Bornean endemic Proboscis monkey *(Nasalis larvatus)*. *Nasalis larvatus* is endemic to the island of Borneo and they favour coastal regions to inland areas.

The Proboscis monkey also known locally as *Monyet Belanda / Bayau* are widespread in Borneo and can be found in Brunei, Indonesia (Kalimantan) and Malaysia (Sabah and Sarawak). These Old World Monkeys which separated from the human lineage 25 million years ago are iconic to Sabah and serve as one of the major tourist attractions in Sabah. Despite being on earth for such a long period, habitat loss has led to the decline in endemic population and listed by the Convention on International Trade in Endangered Species (CITES) as Appendix I.

The alarming rate in the declination of the population had researchers light upon effective measure to control the situation. The rationale of this study is to first help find unique genes in the whole genome of the Proboscis Monkey that could contribute in the conservation of the species and secondly to find genes that could help researchers in deciphering the role and function of genes that might be involved in immunity. By knowing the role and function of these genes, it might provide clues on how the species adapt to certain diseases and also find out if these genes are conserved in the Proboscis monkey and in human as well.

1.2 Significant of Study

Maternally inherited mtDNA is proven to be valuable for investigating intraspecific variation, population structure, phylogeography, and demography in these species. Recent phylogenetic studies divulge that mtDNA sequences are able to offer sufficient resolution for reconstructing a strong phylogeny. These data produce a reliable phylogeny featuring robust support values and most conspicuously contribute to resolution of the phylogeny of *Nasalis* species.

Previous studies have reported genetic variation of Proboscis monkey; there is limited genetic information available regarding this species (Meijaard & Nijman, 2000a). To get an understanding of genetic variation and phylogeny of Proboscis monkey, the mitochondrial DNA of the species from the island of Borneo should be analysed. This species is almost never seen in captivity outside of Borneo, and it is quite rare in zoos as well. While previous studies have lent insights into the taxonomy and phylogeny of the genus, phylogenetic relationships estimates on Proboscis Monkey remain unclear. As such, this study is an attempt to understand the genetic basis adaptation of these critically endangered primates by constructing a highly resolved phylogenetic tree with strong support for relationship between Proboscis monkey and other primates using Maximum likelihood and Neighbour-joining trees.

The major histocompatibility complex (MHC) consists of the most diverse genes identified in vertebrates (Knapp, 2005). Many decades of genetic research revealed that MHC genes play a critical role in the immune response and disease resistance. According to more recent studies, it is suggested that MHC genes influence individual odours used in kin recognition and mate choice as the MHC polymorphism acts as a tool for discriminating among individuals of the same species. For these reasons, the genes of the MHC can provide important perception into evolutionary origins, natural selection, and behavioural ecology in humans and nonhuman primates (Setchell *et al.*, 2010; Zelano *et al.*, 2002). In addition to external fitness related traits, the signal used in MHC based mate choice may also depend on odour which allows distinguishing MHC identities. MHC complexes that are not sustained at the surface of the cell but alternatively released into the